

# Draft Genome Sequence of the Hydrogen- and Ethanol-Producing Bacterium *Clostridium intestinale* Strain URNW

Sadhana Lal,<sup>a</sup> Umesh Ramachandran,<sup>a</sup> Xiangli Zhang,<sup>c</sup> Richard Sparling,<sup>b</sup> David B. Levin<sup>a</sup>

Department of Biosystems Engineering, University of Manitoba, Winnipeg, Manitoba, Canada<sup>a</sup>; Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada<sup>b</sup>; Department of Plant Science, University of Manitoba, Winnipeg, Manitoba, Canada<sup>c</sup>

**Here, we report the draft genome sequence of *Clostridium intestinale* strain URNW, which can convert biomass to useful products such as biofuels (hydrogen or ethanol) and other soluble end products.**

Received 23 September 2013 Accepted 25 September 2013 Published 17 October 2013

**Citation** Lal S, Ramachandran U, Zhang X, Sparling R, Levin DB. 2013. Draft genome sequence of the hydrogen- and ethanol-producing bacterium *Clostridium intestinale* strain URNW. *Genome Announc.* 1(5):e00871-13. doi:10.1128/genomeA.00871-13.

**Copyright** © 2013 Lal et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to David B. Levin, david.levin@umanitoba.ca.

*Clostridium intestinale* strain URNW is a Gram-positive, mesophilic, anaerobic spore-forming bacterium (1). Its optimum growth temperature is 37°C, but it can grow at a range of temperatures from 22°C to a maximum of 45°C and has displayed a doubling time of 1.5 h (1). Nucleotide sequence analyses of *Clostridium intestinale* URNW 16S rRNA genes and chaperonin 60 (*cpn60*) genes revealed a 98% identity with *C. intestinale* ATCC 49213 (1, 2). Phylogenetic analyses based on 16S rDNA and *cpn60* nucleotide sequences indicated that *C. intestinale* URNW clustered with butyrate-producing hydrogen producers, such as *C. intestinale*, *C. acetobutylicum*, *C. perfringens*, *C. butyricum*, and *C. beijerinckii* (1).

The genome of *C. intestinale* strain URNW was sequenced by the McGill University and Genome Quebec Innovation Centre by use of the paired-end and shotgun protocols of the Roche/454s GS-FLX Titanium system and the paired-end protocol of the Illumina HiSeq 2000 platform. The raw reads of the 454 and Illumina were first preprocessed with GS2.6 and Trimmomatic-0.22 (3) for adapter clipping and quality trimming before assembly. After pre-processing, mixed paired-end and unpaired Illumina short reads and 454 long reads were pooled for assembly with the Velvet (4) assembler. Following Velvet assembly, further scaffolding and gap filling were performed with SSPACE v2.0 (5) and GapFiller v 1.11 (6). The final draft genome assembly has coverages of ~600 for Illumina reads and ~28 for 454 reads and contains 39 contigs with a total size of 4,671,178 bp, an *N*<sub>50</sub> contig length of 351,514 nucleotides, and a mean G+C content of 30.2%. The average contig length of is 119,774 bp and the longest contig length is 1,533,141 bp.

The draft genome assembly was submitted to NCBI WGS genome sequencing project and annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The NCBI PGAP annotation GenBank file was uploaded to the web service of the Gene Prediction Improvement Pipeline (GenePRIMP) (7) developed by JGI for analysis of gene anomalies. Based on the report of GenePRIMP anomalies, manual curation was done using Artemis (8). The final annotation of the draft genome sequence of *C. in-*

*testinale* strain URNW has a total of 4,660 genes, including 4,367 protein-coding genes, 182 pseudogenes, 83 tRNAs, and 28 rRNAs.

*C. intestinale* strain URNW protein-coding genes were verified using other *Clostridium* species as reference organisms. Amino acid sequences for each gene product were retrieved from the JGI genome portal (<http://genome.jgi-psf.org/>) and the NCBI database (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>), and sequence alignments against *C. intestinale* strain URNW genes were performed. Key enzymes involved in core metabolism were identified using an *in silico* approach, which revealed that *C. intestinale* strain URNW is a potential candidate for the production of biofuels such as hydrogen or ethanol. All 39 contigs were analyzed by the online software IslandViewer and did not show any genes associated with pathogenicity (9).

**Nucleotide sequence accession numbers.** The genome sequence of *Clostridium intestinale* strain URNW has been deposited at DDBJ/EMBL/GenBank under the accession no. [APJA000000000](https://www.ncbi.nlm.nih.gov/nuclot/APJA000000000). The version described in this paper is the first version, accession no. [APJA010000000](https://www.ncbi.nlm.nih.gov/nuclot/APJA010000000).

## ACKNOWLEDGMENTS

This work was supported by funds from Genome Canada, through the Applied Genomics Research in Bioproducts or Crops (ABC) program for the grant titled “Microbial genomics for biofuels and coproducts from biorefining processes,” and the Province of Manitoba from the Manitoba Research Innovation Fund (MRIF).

## REFERENCES

1. Ramachandran U, Wrana N, Cicek N, Sparling R, Levin DB. 2011. Isolation and characterization of a hydrogen- and ethanol-producing *Clostridium* sp. strain URNW. *Can. J. Microbiol.* 57:236–243.
2. Lee W-K, Fujisawa T, Kawamura S, Kikuiji I, Mitsuoka T. 1989. *Clostridium intestinalis* sp. nov., an aerotolerant species isolated from the feces of cattle and pigs. *Int. J. Syst. Bacteriol.* 39:334–336.
3. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40:W622–W627.
4. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
5. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579.

6. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol.* 13:R56.
7. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* 7:455–457.
8. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28: 464–469.
9. Langille MGI, Brinkman FSL. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25:664–665.